Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles

Achuta Kadambi[1]*,   Refael Whyte[2,1],   Ayush Bhandari[1],   Lee Streeter[2],
Christopher Barsi[1],   Adrian Dorrington[2],   Ramesh Raskar[1]
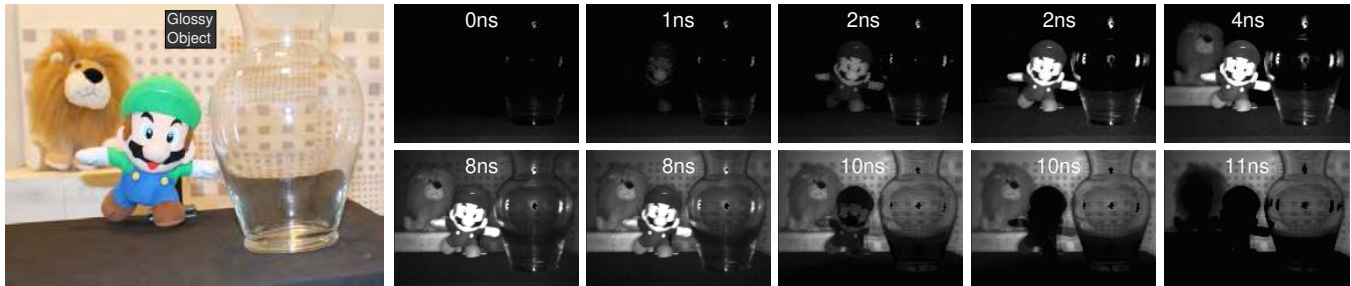[1]Massachusetts Institute of Technology, Boston USA   [2]University of Waikato, Waikato NZ

Figure 1: *Using our custom time of flight camera, we are able to visualize light sweeping over the scene. In this scene, multipath effects can be seen in the glass vase. In the early time-slots, bright spots are formed from the specularities on the glass. Light then sweeps over the other objects on the scene and finally hits the back wall, where it can also be seen through the glass vase (8ns). Light leaves, first from the specularities (8-10ns), then from the stuffed animals. The time slots correspond to the true geometry of the scene (light travels 1 foot in a nanosecond, times are for round-trip). Please see http://media.mit.edu/~achoo/lightsweep for animated light sweep movies.*
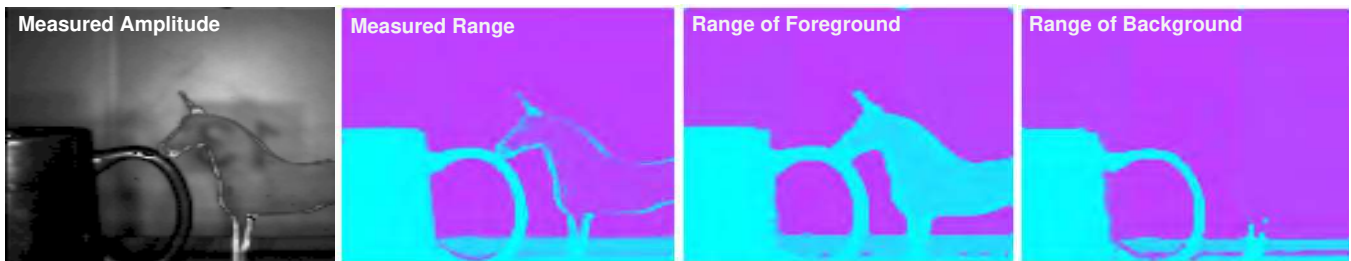

Figure 2: *Recovering depth of transparent objects is a hard problem in general and has yet to be solved for Time of Flight cameras. A glass unicorn is placed in a scene with a wall behind (left). A regular time of flight camera fails to resolve the correct depth of the unicorn (center-left). By using our multipath algorithm, we are able to obtain the depth of foreground (center-right) or of background (right).*

Abstract

Time of flight cameras produce real-time range maps at a relatively low cost using continuous wave amplitude modulation and demodulation. However, they are geared to measure range (or phase) for a single reflected bounce of light and suffer from systematic errors due to multipath interference.

We re-purpose the conventional time of flight device for a new goal: to recover per-pixel sparse time profiles expressed as a sequence of impulses. With this modification, we show that we can not only address multipath interference but also enable new applications such as recovering depth of near-transparent surfaces, looking through diffusers and creating time-profile movies of sweeping light.

Our key idea is to formulate the forward amplitude modulated light propagation as a convolution with custom codes, record samples by introducing a simple sequence of electronic time delays, and perform sparse deconvolution to recover sequences of Diracs that correspond to multipath returns. Applications to computer vision include ranging of near-transparent objects and subsurface imaging through diffusers. Our low cost prototype may lead to new insights regarding forward and inverse problems in light transport.

---

*achoo@mit.edu

Links: ◆DL 📕PDF

1. Introduction

Commercial time of flight (ToF) systems achieve ranging by amplitude modulation of a continuous wave. While ToF cameras provide a single optical path length (range or depth) value per pixel the scene may actually consist of multiple depths, e.g., a transparency in front of a wall. We refer to this as a mixed pixel. Our goal is to recover the sequence of optical path lengths involved in light reaching each pixel expressed as a time profile. To overcome the mixed pixel problem and to enable new functionality, we repurpose the device and change the computation in two ways. First, we emit a custom code and record a sequence of demodulated values using successive

electronic delays. Second, we use a sparse deconvolution procedure to recover a sequences of Diracs in the time profile corresponding to the sequence of path lengths to multipath combinations.

1.1. Contributions

Our key contribution:

- The idea of using a single-frequency, coded illumination ToF camera to turn time profile recovery into a well-conditioned sparse deconvolution problem.

Secondary technical contributions:

- Evaluation of different modulation codes and sparse programs.
- A tailored sparse deconvolution model using a proximity-based approach to matching pursuit.

These technical contributions lead to four applications for time of flight cameras:

1. Light Sweep Imaging.[1]
2. Obtaining depth of translucent objects.
3. Looking through diffuse media.
4. Increasing the accuracy of depth maps by solving and correcting for multipath reflections.

In addition, we provide quantitative evaluations of our methods, suggest directions for future work, and discuss the benefits and limitations of our technique in the context of existing literature.

Scope: In this paper, we construct a low-cost prototype camera from a bare sensor; however, our method can be implemented on commercial time of flight cameras by reconfiguring the proprietary, on-board FPGA software. Constructing a similar prototype to our camera would cost 500-800 dollars. In this paper, we consider recovery of a discretized time profile, which means that our technique cannot resolve closely spaced optical paths that mix into 1 time slot, such as the numerous interreflections from a corner.

2. Related Work

Time Profile Imaging dates back to the work of Abramson in the late 70's. Abramson's technique, so-called "light in flight", utilized holographic recordings of a scene to reconstruct the wavefront of light [Abramson 1980]. In 2011 the nascent field of *femto-photography* was introduced in the vision and graphics community. Key papers include: reflectance capture using ultrafast imaging [Naik et al. 2011], frequency analysis of transient light transport [Wu et al. 2012], looking around corners [Velten et al. 2012], and *femto-photography* [Velten et al. 2013]. While a photonic mixer device has been proposed for low-budget transient imaging [Heide et al. 2013], this technique requires data at hundreds of frequencies of modulation, to recover a time profile. In contrast, we implement a radically simple single-frequency, time shifted capture that addresses sparsity in optical path lengths as well as multipath.

Multipath Interference occurs when multiple light-paths hit the ToF sensor at the same pixel. This results in a measured range that is a non-linear mixture of the incoming light paths. Dorrington demonstrated the seminal method to resolve multiple on commercial ToF cameras by using multi-frequency measurements [Dorrington et al. 2011; Bhandari et al. 2013]. Godbaz continued this

---

[1]Light Sweep Imaging is a visualization of the per-pixel time profile at 70 ps time resolution achieved by sweeping through the time profile solved with the Tikhonov program. The output is analogous to a temporally encoded depth map, where the multipath problem has been solved. Figure 1 illustrates the distinction where a proper time profile movie handles mixing on the pixel-level from the vase and wall behind.

work by finding a closed-form mathematical model method to mitigate multifrequency multipath [Godbaz et al. 2012]. The work by Heide et al. investigates multipath in the context of global illumination, which was proposed in [Raskar and Davis 2008]. A number of methods of separating out multiple returns based on the correlation waveform shape were explored by Godbaz. Specifically, this included investigations into sparse spike train deconvolution via gradient descent, the Levy-Fullagar algorithm, and a waveform shape fitting model to accurately model. Godbaz concluded that the correlation function in standard AMCW is not designed for harmonic content, which limits all current methods [Godbaz et al. 2013]. Previous attempts only address multipath in the context of mitigating range errors. In this paper, we recover multipath returns by demultiplexing at a *single* modulation frequency.

Custom Codes have been recently proposed within the time of flight community to reduce interference from an array of multiple time of flight cameras. While the standard code is a simple square pulse, Buttgen et al. and Whyte et al. have proposed custom codes in the specific context of simultaneous camera operation: based on these research lines, a patent was granted to Canesta Inc [Buttgen and Seitz 2008; Whyte et al. 2010; Bamji et al. 2008]. However, custom camera codes have not been explored for the case of single camera operation. Although, coding in time with carefully chosen binary sequences is used for motion deblurring in conventional cameras [Raskar et al. 2006], for ToF cameras, we show that they can be used to resolve multipath returns.

Sparse Deconvolution was first introduced in the context of seismic imaging. While Weiner deconvolution was the prevalent approach until an $\|\cdot\|_{\ell_1}$ penalty term (where $\|\mathbf{x}\|_{\ell_1} = \sum_k |x_k|$) in context of sparsity inducing decovolution was introduced by Clearbout and Muir [Claerbout and F 1973] in 1970s. Taylor et al. [Taylor et al. 1979] worked on a variation to seek the solution to the problem $\mathcal{J}_\lambda (\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_{\ell_1} + \lambda \|\mathbf{x}\|_{\ell_1}$. [Santosa and Symes 1986] recast the least–squares deconvolution problem with sparse penalty term, $J_\lambda (\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_{\ell_2}^2 + \lambda \|\mathbf{x}\|_{\ell_1}$. Since then a number of modifications have been proposed for both the cost function $\mathcal{J}_\lambda$, for example see [O'Brien et al. 1994] as well as the problem of accelerating the optimization problem with ℓ_1 penalty term [Darche 1989; Daubechies et al. 2004] with Lasso and Basis–pursuit being notable examples. These problems are now discussed under the general theme of sparsity and compressed sensing [CSm 2008].

Depth Imaging of Translucent Objects is an active problem in range imaging. In the context of structured light, Narasimhan et al. were able to obtain the range of objects in scattering media by using five separate illumination locations [Narasimhan et al. 2005]. For stereo imaging, Tsin et al. were able to show that their stereo matching algorithm was able to range through translucent sheets [Tsin et al. 2006]. A recent paper from [Gupta et al. 2013] delivered a practical approach to 3D scanning in the presence of different scattering effects. Despite the increasing popularity of time of flight technology, ranging of transparent objects remains an open problem for the time of flight community.

Subsurface Rendering and Imaging has been a cornerstone in the computer graphics and computer vision communities. It is well known that accurate modelling of global illumination is integral to realistic renderings [Jensen et al. 2001]. Approaches for imaging through a diffuser have included time-resolved reconstruction [Naik et al. 2013] and a sparse, spatial coding framework [Kadambi et al. 2013]. In this paper, we look through a diffuser by coding in time with ToF camera hardware.
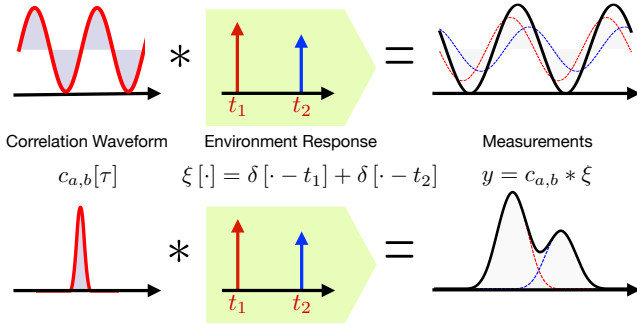
Figure 3: *Why use a Custom, Coded Camera? (top row) When using a conventional time of flight camera the correlation waveform is sinusoidal (red curve). When this is convolved with the environment response, the resulting measurement (black curve) is also sinusoidal. This creates a problem in unicity. (bottom row) When using custom codes, the correlation waveform shows a distinct peak that is non-bandlimited. Therefore, when convolved with the environment, the output measurement has 2 distinct peaks. The diracs can be recovered by a deconvolution, specifically, a sparse deconvolution. This is the crux of our paper.*

3. Time of Flight Preliminaries

3.1. Overview of Conventional Time of Flight

A ToF camera is, for the most part a regular camera, with a time-coded illumination circuit. Concretely, in common, commercial implementations, an LED is strobed with a high-frequency square pulse (often 30-50 MHz). Light hits an object in the scene and then returns to the camera. Sampling the optical signal in time reveals a shifted version of the original signal, where the amount of shift encodes the distance that light has travelled. One approach to calculating the shift is to simply cross-correlate the reference waveform with the measured optical signal and find the peak.

3.2. Notation

Throughout this discussion, we use $f(\cdot)$ for functions with continuous argument and $f[\cdot]$ as their discrete counterparts. Also, a major part of our discussion relies on the definition of cross–correlation between continuous functions and discrete sequences. Given two functions $a(t)$ and $b(t)$, we define the cross–correlation as

$$c_{a,b}(\tau) = \lim_{\Delta \to \infty} \frac{1}{2\Delta} \int_{-\Delta}^{+\Delta} a^*(t) b(t+\tau) dt \Leftrightarrow (a \otimes b)(t) \; \forall t \in R$$

where a^* denotes complex–conjugate of a and $\otimes$ denotes the cross–correlation operator. Cross–correlation is related to the convolution operator by:

$$(a \otimes b)(t) = (\bar{a}^* * b)(t)$$

where $\bar{a}(t) = a(-t)$ and '$*$' denotes the linear convolution operator.

The definition of cross–correlation leads to a natural extension for discrete sequences:

$$c_{a,b}[\tau] = \frac{1}{K} \sum_{k=0}^{K-1} a^*[k] b[\tau+k] \Leftrightarrow (a \otimes b)[\tau]$$

where $a[k] = a(kT)$ and $b[k] = b(kT), \forall k \in Z$ and for some sampling step $T > 0$.

A ToF camera uses an illumination control waveform $i_\omega(t)$ with a modulation frequency ω to strobe the light source. In practice, it is

often the case that the illumination waveform is a periodic function such that:

$$i_\omega(t + T_0) = i_\omega(t)$$

where T_0 is the time–period of repetition. Since we use a homo-dyne setting, for the sake of notational simplicity, we will drop the subscript ω and use $i = i_\omega$. The TOF camera measurements are obtained by computing $c_{m,r}(\tau)$ where,

- $m(t)$: optical signal from the light source, and,
- $r(t)$: reference signal.

In typical implementations the illumination control signal and reference signal are the same, that is, $i(t) = r(t)$. The phase which is encoded in the shift $\tau^\star = \arg \max_\tau c_{m,r}[\tau]$, can be obtained a number of ways.

In commercial implementations, for example the PMD or the Mesa TOF cameras, 2 to 4 samples of the correlation function $c_{m,r}[\tau]$ suffice for the computation of the phase. For many modulation functions, a sample on the rising edge and another on the falling edge are sufficient to find the peak. Another technique for computing the phase involves oversampling of the correlation function. There on, it is possible to interpolate and analyse the Fourier spectrum or simply interpolate the peak directly. The oversampling case is germane to the problem of multipath as the correlation function (for a custom code) can become distorted. The final calculation from phase to distance is a straightforward linear calculation. For a Photonic Mixer Device that uses square wave modulation and a sinusoidal form for $c_{m,r}[\tau]$, the conversion is simply:

$$d = \frac{c\phi}{4\pi f_\omega}, \quad c = 3 \times 10^8 \text{ m/s} .$$

It is important to note that time of flight cameras can theoretically operate at different modulation frequencies, which means that the distance is constant at different modulation frequencies and thus the ratio $\frac{\phi}{f_\omega}$ is constant, that is, doubling the modulation frequency will double the phase for a single-path scene.

3.3. Custom Codes

Conventional implementations use a sinusoidal correlation function. This approach works for conventional range imaging, but cannot deal with multipath objects. In figure 3, we see that the two reflected sine waves from the red and blue objects add to produce a sinusoidal measurement (black curve). The phase of this measured sinusoid is in-between the phases of the component red and blue sinusoids, which creates a problem of unicity. Concretely, it is unclear whether two component sine waves are really in the environment, or if only one component exists (with the mixed phase).

We now turn to using custom codes. Figure 3 illustrates that it is desirable to change the correlation waveform to avoid problems with unicity. This can be done by selecting appropriate binary sequences for $r(t)$ and $i(t)$, which we detail in Section 6 and Figure 5. In Section 4 we show that the code selection also ties in with the conditioning of our inverse problem.

3.4. Benefits of Single Frequency

An alternate approach introduced by the Waikato Range Imaging group is to acquire range maps at different modulation frequencies and then solve a fitting problem to resolve multipath—this is also the method used by Heide et al. Unfortunately, the problem of exponential fitting is known to be ill-conditioned and the implementation is often challenging—it is time consuming, requires additional hardware for multi-frequency, and most important, the frequency
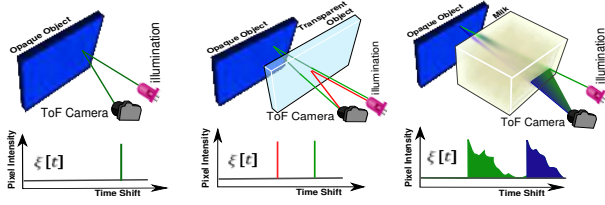
Figure 4: *The environment profile function $\xi[t]$ is simply a discretized time profile. For the case of one opaque wall, the environment profile will appear as a single spike. For a transparency in front of a wall, the environment profile appears as a pair of spikes. In the case of scattering media, the time profile may not be sparse. For most applications of time of flight a sparse formulation of $\xi[t]$ is desired.*

response calibration varies from shot to shot. In this paper, multipath recovery is performed using only a single frequency ToF camera. Such a camera can be built with reprogramming of the FPGA on PMD and Mesa cameras.

4. Coded Deconvolution

4.1. Forward Model of Environment Convolution

We start by relating the optically measured signal $m[t]$ to the discrete illumination control signal $i[t]$:

$$m[t] = (i * \varphi * \xi)[t]. \tag{1}$$

Here, the illumination signal $i[t]$ is first convolved with a low–pass filter $\varphi[t]$. This represents a smoothing due to the rise/fall time of the electronics. Subsequent convolution with an environment response $\xi[t]$ returns the optical measurement $m[t]$.

The function $\xi[t]$ is a scene-dependent time profile function (Figure 4). For a single opaque object, $\xi[t]$ appears as a Kronecker Delta function: $\delta[t - \phi]$ where ϕ represents the sample shift that encodes path-length and object depth. In the multipath case, without scattering, the environment function represents a summation of discrete Dirac functions:

$$\xi[t] = \sum_{k=1}^{K-1} \alpha_k \delta[t - t_k],$$

where $\{\alpha_k, \tau_k\}_{k=0}^{K-1}$ denotes amplitude scaling and phases, respectively.

We now turn to a definition of the measured cross-correlation function in the presence of the environment function:

$$
\begin{aligned}
c_{r,i*\varphi*\xi}[\tau] &= (r \otimes (i * \varphi * \xi))[\tau] \\
&= \underbrace{(r \otimes i)}_{\zeta[t]} * \varphi * \underbrace{\sum_{k=0}^{K-1} \alpha_k \delta[\cdot - t_k]}_{\text{Sparse Environment Response}} \\
&= \zeta * \varphi * \sum_{k=0}^{K-1} \alpha_k \delta[\cdot - t_k].
\end{aligned}
\tag{2}
$$

This is the key equation in our forward model where our measurements $c_{r,i*\varphi*\xi}[\tau]$ are the cross-correlations in presence of an unknown, parametric environment response, $\xi[t]$. In this paper, we have written the measurement as a convolution between the environment and the *deterministic* kernel, $\zeta[t] = (r \otimes i)[t]$ and the low pass filter, φ.

Conditioning the Problem Note that Equation 2 can be developed as:

$$\underbrace{(\zeta * \varphi)}_{h} * \xi[t] = (h * \xi)[t] \tag{3}$$

where $h[t]$ is the convolution kernel resulting from low–pass filtering of ζ.

In vector–matrix form the convolution is simply a circulant Toeplitz matrix acting on a vector:

$$y = \underbrace{(h * \xi)}_{\mathbf{Hx}}[t] \Leftrightarrow \underbrace{\mathbf{H}^{d \times d}}_{\text{Toeplitz}} : \mathbf{x}^{d \times 1} \mapsto \mathbf{y}^{d \times 1} = \mathbf{Hx} \tag{4}$$

where $\mathbf{y} \in R^d$ is measurement vector which amounts to the sampled version of the correlation function where d represents the number of samples. The convolution matrix $\mathbf{H} \in R^{d \times d}$ is a circulant Toeplitz matrix, where each column is a sample–shift of $h[t]$. Since h implicitly containts a low–pass filter φ. Finally, the vector $\mathbf{x} \in R^d$ is the vector corresponding to the environment $\xi[t]$,

$$\mathbf{x} = [\xi[0], \xi[1], \ldots, \xi[d-1]]^\top.$$

Given $\mathbf{y}$, since we are interested in parameters of ξ, the key requirements on the convolution matrix $\mathbf{H}$ is that its inverse should be well defined. Equation 4 is a classic linear system. Provided that $\mathbf{H}$ is well conditioned, it can be inverted in context of linear inverse problems. Since $\mathbf{H}$ has a Toeplitz structure, it is digonalized by the Fourier matrix and the eigen–values correspond to the spectral components of h.

Controlling the condition number of $\mathbf{H}$ amounts to minimizing the ratio of highest to lowest Fourier coefficients of $h[t]$ or eigen–values of $\mathbf{H}$. This is the premise for using binary sequences with a broadand frequency response. Figure 5 outlines several common codes as well as their spectrums.

In this paper, we assess different code strategies in section 6.

4.2. Sparse Formulation

Since $\xi[t]$ is completely characterized by $\{\alpha_k, \tau_k\}_{k=0}^{K-1}$, in the multipath case, our goal is to estimate these parameters. For given set of measurements y,

$$y[\tau] = \sum_{k=0}^{K-1} \alpha_k h[\tau - t_k] \Leftrightarrow \mathbf{y} = \mathbf{Hx}$$

and knowledge of h, the problem of estimating ξ boils down to,

$$\arg \min_{\{\alpha_k, t_k\}} \left| \sum_{k=0}^{K-1} y[t] - \sum_{k=0}^{K-1} \alpha_k h[t - t_k] \right|^2.$$

There are many classic techniques to solve this problem in time or in frequency domain, including a pseudoinverse or even Tikhonov regularization. However, since we know that ξ is a K–sparse signal, in this paper we begin with a sparsity promoting optimization scheme. The problem falls into the deconvolution framework mainly because of the low–pass nature of h or the smearing effect of $\mathbf{H}$. In this context, the sparse deconvolution problem results in the following problem:

$$\arg \min_{\mathbf{x}} ||\mathbf{Hx} - \mathbf{y}||_2^2 \quad \text{such that} \quad ||\mathbf{x}||_0 \leq K,$$

where $||\mathbf{x}||_0$ is the number of non-zero entries in $\mathbf{x}$. Due to non–convexity of $||\mathbf{x}||_0$ and mathematical technicalities, this problem is intractable in practice. However, a version of the same which incoporated convex relaxation can be cast as:

$$\arg \min_{\mathbf{x}} ||\mathbf{Hx} - \mathbf{y}||_2^2 \quad \text{such that} \quad ||\mathbf{x}||_1 \leq K,$$
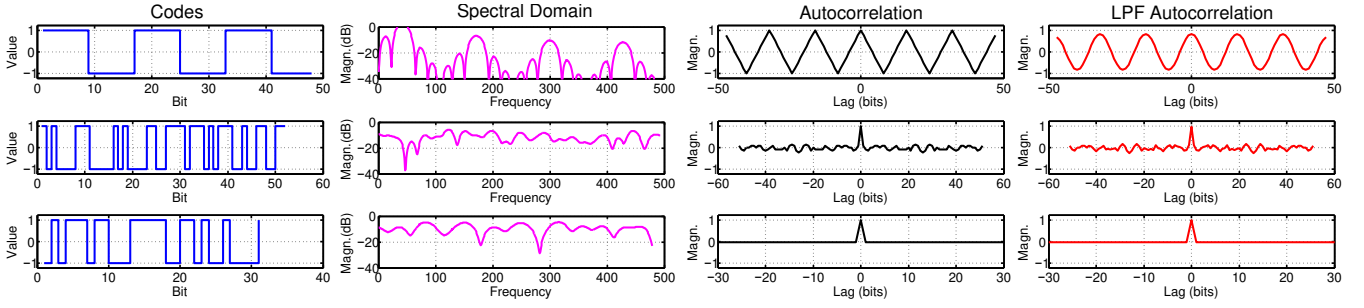
Figure 5: *Evaluating custom codes. We compare three different codes: conventional square, broadband code from [Raskar et al. 2006], and our m-sequence. The spectrum of the codes (magenta) affects the condition number in our linear inverse problem. The autocorrelation function (black) is simply the autocorrelation of the bit sequences shown in blue. In the context of physical constraints, a low pass filter smoothens out the response of the correlation function to provide the measured autocorrelation shown in red. Note that the spectrum of the magenta curve contains nulls (please zoom in PDF).*

where $\|\mathbf{x}\|_1 = \sum_k |x_k|$ is the ℓ_1–norm. This is commonly known as the LASSO problem. Several efficient solvers exist for this problem. We use `SPGL1` and `CVX` for the same. An alternative approach to utilizes a modified version of the greedy, orthogonal matching pursuit algorithm (OMP) that approximates the ℓ_0 problem. In particular, we propose two modifications to the OMP formulation:

1. Non-negativity Constraints, and,

2. Proximity Constraints

Non-negativity requires two modifications to OMP: (a) when searching for the next atom, consider only positive projections or inner products, and (b) when updating the residual error, use a solver to impose a positivity constraint on the coefficients. Detailed analysis including convergence bounds can be found in [Bruckstein et al. 2008].

Our second proposition which involves proximity constraints is similarly simple to incorporate. For the first projection, we allow OMP to proceed without modifications. After the first atom has been computed, we know that the subsequent atom must be in proximity to the leading atom in the sense that the columns are near one another in the matrix $\mathbf{H}$. In practice, this involves enforcing a Gaussian penalty on the residual error. This can be formulated as a maximum a posteriori (MAP) estimation problem:

$$\arg \max_{\mathbf{x}} p(x|y) \propto \arg \max_{\mathbf{x}} \underbrace{p(y|x)}_{\text{likelihood}} \underbrace{p(x)}_{\text{prior}}$$

where $p(y|x)$ is the likelihood which is a functional form of the combinatorial projection onto the dictionary, and $p(x)$ is a prior, modelled as,

$$p(x) \in \mathcal{N}(x; \mu, \sigma^2) \text{ where } \mu = x_{K=1}$$

where $\mathcal{N}$ is the usual Normal Distribution with mean and variance μ and σ^2, respectively. Here, $x_{K=1}$ represents the column index of the first atom.

In our case, the prior is a physics–inspired heuristic that can be carefully chosen in the context of binary codes and by extension the knowledge of our convolution kernel.

4.3. Deconvolution for Time Profile Movies

In a transient movie each pixel can be represented as a time profile vector which encodes intensity as a function of time. We recognize that for a time of flight camera, the analogue is a phase profile vector, or in other words the environment function $\xi[t]$. In the previous case, we considered $\xi[t]$ to be a sparse function resulting from a few objects at finite depths. However, in the case of global illumination
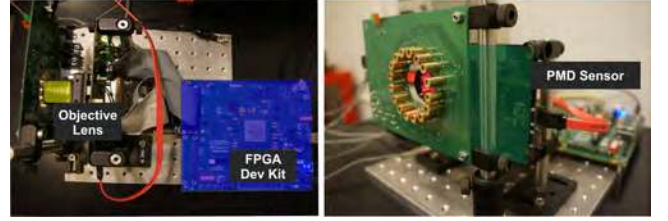

Figure 6: *Our prototype implementation uses a Stratix FPGA with high-frequency laser diodes interfaced with a PMD19k-2 sensor. Similar functionality can be obtained on some commercial ToF cameras with reconfiguration of the FPGA programs.*

and scattering, the environment response is a non-sparse function (see Figure 4). To create a transient movie in the context of global illumination we use Tikhonov regularization for our deconvolution problem. Specifically, we solve the problem in the framework of Hodrick–Prescott filtering [Hodrick and Prescott 1997] which can be thought of as Tikhonov regularization with a smoothness prior:

$$\arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Hx}\|_2^2 + \lambda \|\mathbf{Dx}\|_2^2,$$

where $\mathbf{D}$ is a second order difference matrix with circulant Toeplitz structure and λ is the smoothing parameter. We use Generalized Cross-Validation to select the optimal value of λ.

5. Implementation

Because we use a single modulation frequency, the hardware protoype requires only a reconfiguration of the on-board FPGA located on commercial ToF cameras. However, on such cameras, the FPGA is surface mounted and the HDL code is proprietary.

Therefore, we validate our technique using a prototype time of flight camera designed to send custom codes at arbitrary shifts (see Figure 6). For the actual sensor, we use the PMD19k-2 which has a pixel array size of 160×120. This sensor is controlled by a Stratix III FPGA operated at a clock frequency of 1800 MHz. For illumination we use Sony SLD1239JL-54 laser diodes that are stable at the modulation frequency we use (50 MHz). The analog pixel values converted to 16bit unsigned values by an ADC during the pixel array readout process. A photo of the assembled camera is shown in figure 6. For further details please refer to [Whyte et al. 2010] and [Carnegie et al. 2011] for the reference design.

Time Resolution: The Stratix III FPGA allows for rapid sweeps between the reference and illumination signal. In our implementation, the modulation signals are generated on the phase lock loop
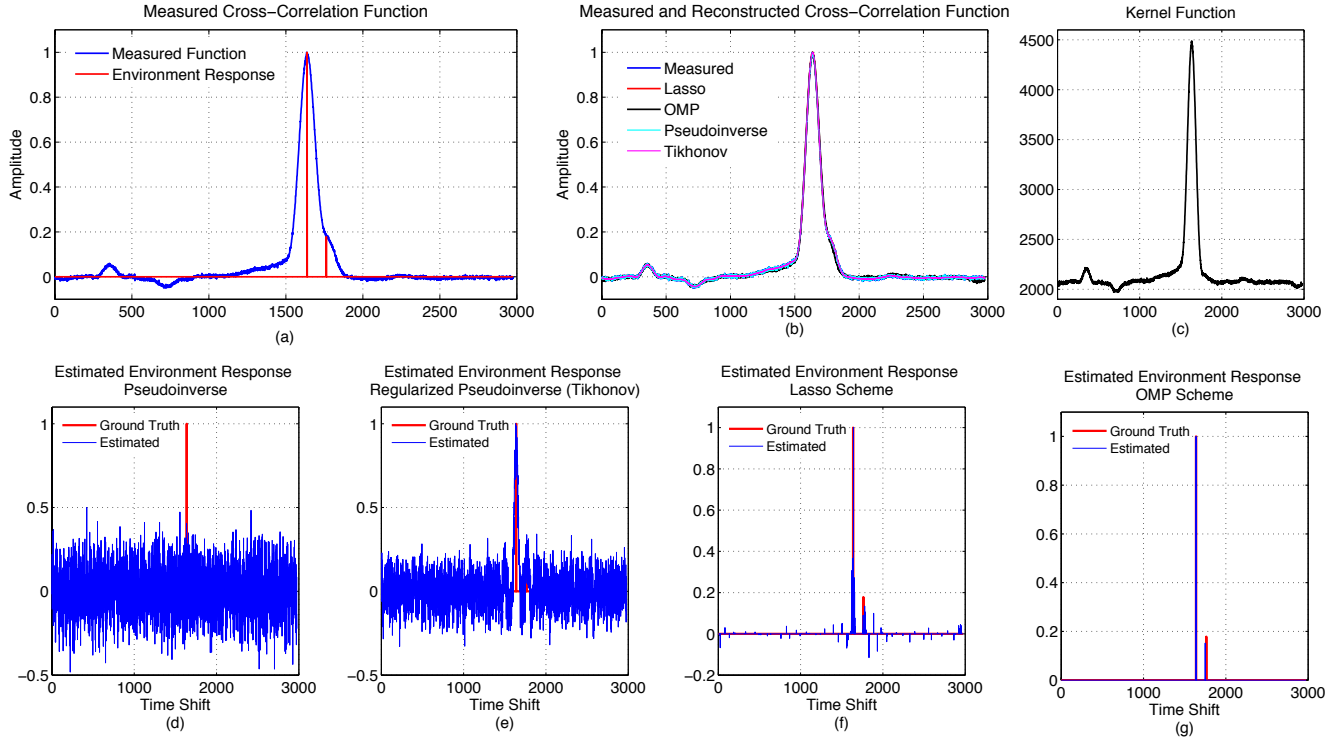
Figure 7: *Comparing different deconvolution algorithms using a mixed pixel on the unicorn dataset (Figure 2). (a) The measured cross correlation function is a convolution between the two diracs and the kernel in (c). (b) The measured and reconstructed cross-correlation function for various classes of algorithms. (d) The deconvolution using a simple pseudoinverse when compared to the ground truth in red. (e) Tikhonov regularization is close to the peak although the solution is not sparse. (f) The Lasso approach hits the peaks but must be tuned carefully to select the right amount of sparsity (g) Finally, our proposed approach is a slightly modified variant of Orthogonal Matching Pursuit. It is able to find the Diracs. This result is from real data.*

(PLL) inside the Stratix III FPGA with a configurable phase and frequency from a voltage controlled oscillator, which we operate at the maximum 1800 MHz. The theoretical, best-case time resolution, is calculated at **69.4 ps** from the hardware specs—please see the supplementary website. From a sampling perspective, this limit describes the spacing between two samples on the correlation waveform.

Future Hardware: The latest Kintex 7 FPGA from Xilinx supports frequencies up to 2133MHz, which would theoretically allow for a time resolution of 58.6ps. It is expected that newer generations of FPGA technology will support increasingly high oscillation frequencies, improving the time resolution of our method.

6. Assessment

6.1. Custom Codes

We now turn to the selection of codes for $r(t)$ and $i(t)$. For simplicity, we consider symmetric coding strategies—we pick the same code for $r(t)$ and $i(t)$. Because the smearing matrix in equation 4 is Toeplitz, its eigenvalues correspond to spectral amplitudes. Since the condition number relates the maximal and minimal eigenvalues, a low condition number in this context corresponds to a broadband spectrum.

Square Codes are used in the typical commercial implementations and lead to a sinusoidal correlation function. This is a double-edged sword. While sinusoidal correlation functions allow a neat parametric method to estimate the phase—one only needs 3 samples to parametrize a sinusoid—they lead to problems of unicity

and are thus not suitable for multipath scenarios (Figure 3).

Delta Codes seem promising for deconvolution as their spectrum is broadband. However, aside from the obvious issue of SNR, it is not possible to generate a true Delta code in hardware. The best you can do is a narrow box function. In Fourier domain this is a sinc function with characteristic nulls which makes the problem poorly conditioned.

Maximum-length sequences are in the class of pseudorandom binary sequences (PN-sequences). PN-sequences are deterministic, yet have a flat spectrum typical of random codes. In particular, the m-sequence is generated recursively from primitive polynomials. The main advantage of m-sequences is that they are equipped with desirable autocorrelation properties. Concretely, let us consider an m-sequence stored in vector z with a period of P:

$$a_{[z,z]}(k) \Leftrightarrow \sum_i z_i \bar{z}_{i-k} = \left\{ \begin{array}{cc} 1 & k = 0 \\ \frac{1}{P} & 0 < k < P - 1 \end{array} \right\} \quad (5)$$

where $a_{[z,z]}$ defines the autocorrelation operator. As the period length P increases the autocorrelation approaches an impulse function, which has an ideal broadband spectral response. As a bonus, m-sequences are easy to generate, deterministic, and spectrally flat.

6.1.1. Simulations

In Figure 5 we show three different codes along with their spectrums, autocorrelation, and "measured autocorrelation" (after the low-pass operator). the spectra of the square code has many nulls, leading to an ill-conditioned problem. Moreover, the autocorrelation of a square code (black curve) smoothens into a sinusoidal

correlation function, which brings the unicity problem from Figure 3 into context. In contrast, the Broadband code from [Raskar et al. 2006], has been optimized to have a very low condition number and flat frequency response.

To summarize: based on the spectrum, either the broadband or m-sequence codes lead to a potentially well-conditioned inverse problem for equation 4. While the Broadband code does have a slightly lower condition number—it has been optimized in that aspect—the m-sequence offers two critical advantages: (i) the code length is easy to adjust and (ii) the autocorrelation function is nearly zero outside of the peak (Equation 5). The length of the m-sequence is critical: too short of a sequence and the autocorrelation will be high outside the peak and too long of a code leads to a longer acquisition time. The code we used is a m-sequence of length 31 (m=5).[2,3]

6.2. Assessing Sparse Programs

In Figure 7 we outline various deconvolution strategies for a mixed pixel on the unicorn in Figure 2. We expect two returns in this scene—one from the surface of the near-transparent unicorn and one from the wall 2 meters behind. From Figure 4 it would seem that two Dirac functions would form a reasonable time profile that convolve with a kernel to provide the measurement. The ground truth is shown as red dirac deltas in Figure 7d-g. The modified variant of orthogonal matching pursuit that we propose seems to perform the best in the sense of sparsity, while Lasso seems to approximate the peaks and amplitudes well. Whichever method is chosen, it is clear that all schemes, including a naive pseudoinverse, lead to a reasonable reconstruction of $\hat{y}$, substantiating our belief that data fidelity combined with sparsity is appropriate for our context (Figure 7b).

7. Results

7.1. Applications

1. Time Profile Imaging: Please see the supplemental website for movie versions. In Figure 1 we see light sweeping first over the vase, then mario, then the lion, and finally to the wall behind. The key idea is that we solve for multipath effects, e.g., the interactions of the translucent vase and back wall. In Figure 8 we see light first sweeping over the teddy bear in the scene, and at later time slots, over its reflection in the mirror.

Now we consider global illumination (due to internal scattering). In Figure 2 a transparent acrylic unicorn (thickness 5 mm) is placed 10 centimeters away from the co-located camera and light source. Approximately 2 meters behind the unicorn is an opaque white wall. We expect two returns from the unicorn—a direct reflection off the acrylic unicorn and a reflection from the back wall passing through the unicorn. The first frame of Figure 9 is acquired at 0.1 ns, where we begin to see light washing over the unicorn. Step forward 200 picoseconds, and we see a similar looking frame, which represents internal reflections. We verify that intensities at this frame exclude surface reflections from the acrylic by observing that the leg, which was specular at 0.1 nanoseconds (Figure 9), has now decreased in intensity. In summary, our technique seems to be able to distinguish direct and global illumination by solving the multipath deconvolution problem. This experiment underscores the link between multipath resolution and time profile imaging (Figure 4).

We turn to physically validating the time resolution of our prototype. In Figure 10 we create a time profile movie of the angled
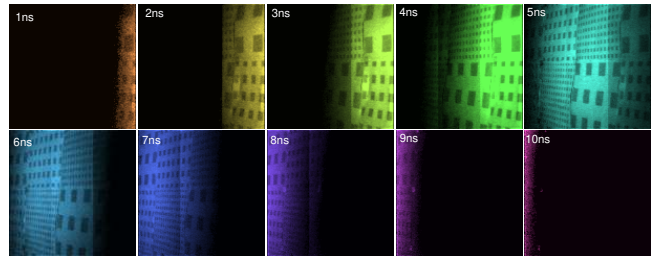
[2]The specific m-sequence: 0101110110001111100110100100001.

[3]See http://media.mit.edu/~achoo/lightsweep/ to generate your own m-sequences.

Figure 10: *Sample frames of Light Sweep Imaging on the angled wall from Figure 11. Since the laser is orthogonal to the normal of the wall, light seems to sweep across the wall. In the best case scenario we are able to obtain a time resolution on the order of 700-1000 picoseconds. Colors represent frames at different times according to the rainbow coloring scheme. The first frame occurs at 1 ns after light has entered the scene and subsequent frames are sampled at every nanosecond. Since light travels 30 cm in a nanosecond we use the geometry of the scene to verify our imaging modality.*
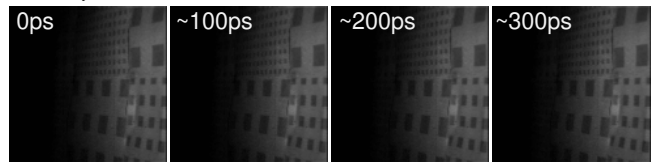

Figure 11: *Quantifying the time resolution of our setup on an angled wall with printed checkers. From the knowledge of the geometry relating the camera, light source and wall as well as the checker size, we calculate the a 6cm optical path length across the large checkers. We show 4 successive frames at approximately 100 picosecond intervals. Within the four frames, light has moved one checker, suggesting that our practical time resolution is approximately on the order of 2 cm or 700-1000 picoseconds. This agrees with our theoretical calculations in Section 5.*

wall. In this setup, the laser beam is nearly orthogonal to the normal of the wall, i.e,. the light strikes nearly parallel along the wall. Using this fact and accounting for the camera position and checker size, we have a calibrated dataset to measure our time resolution: light propagation across the largest square represents 6 cm of round trip light travel (the square is approximately 3 cm). In Figure 11, we show four consecutive frames of the recovered time profile. It takes approximately three frames for the wavefront to propagate along one checker, which suggests we can resolve light paths down to 2 cm. This agrees with the theoretical best case calculated in Section 5. We have verified that the time profile frames correspond with the geometry of the scene.

2. Ranging of Transparent Objects For the unicorn scene, Figure 2 depicts the amplitude and range images that a conventional time of flight camera measures. Because the unicorn is made of acrylic and near-transparent, the time of flight camera measures an incorrect depth for the body of the unicorn. By using sparse deconvolution and solving the multipath problem, we are able to select whether we want to obtain the depth of the unicorn or the wall behind. Our method generalizes to more than two objects. In Figure 13 we show a mixed pixel of 3 different path-lengths. In practice, 2 path lengths are more common and intuitive and are the focus for our applications. Of course, this method is limited by the relative amplitude of foreground and background objects—at specularities there is very little light coming from the back wall.

3. Looking Through Diffusing Material In Figure 12, a diffuser is placed in front of a wall containing printed text. With the regular
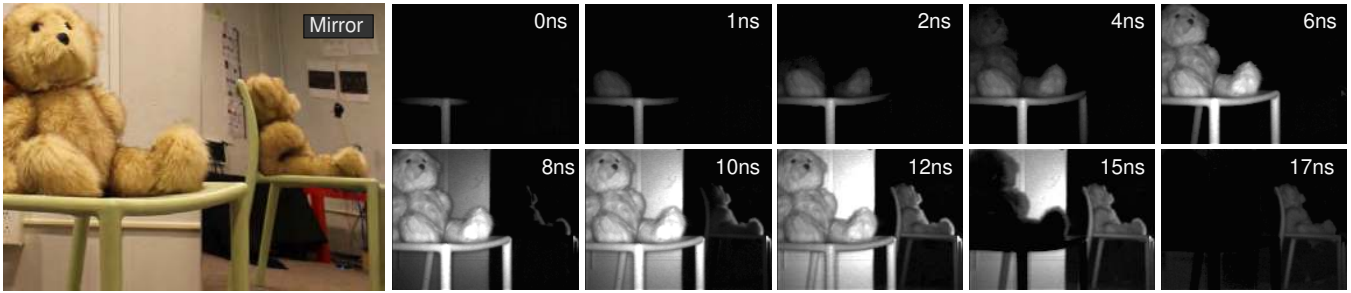
Figure 8: *A light sweep scene where a teddy bear is sitting on a chair with a mirror placed in the scene. We visualize light sweeping, first over the teddy bear (0-6 ns), then to its mirror reflections (8-17 ns). Light dims from the real teddy bear (15 ns) until finally only the reflection persists (17ns). Please see the webpage for the light sweep movie.*
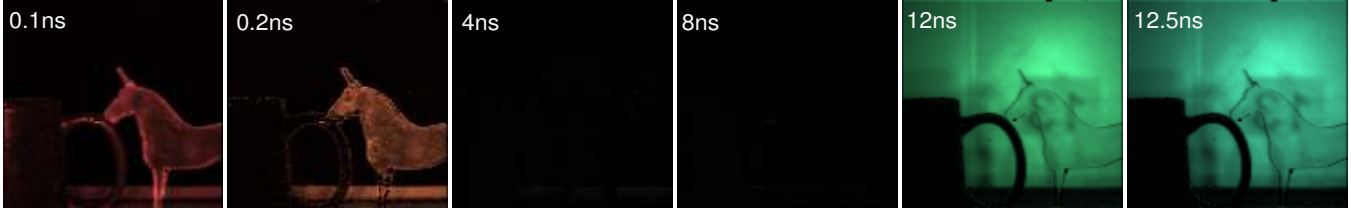

Figure 9: *A light sweep scene with a near-transparent acrylic unicorn 2 meters in front of a wall. In particular, note the large gap between light sweeping over the unicorn and the back wall. The number "13" printed on the back wall, is only visible at later time-slots, while the body of the unicorn is opaque at early time slots. Between the first two frames, the specularities have dissapeared and only the global illumination of the unicorn persists.*
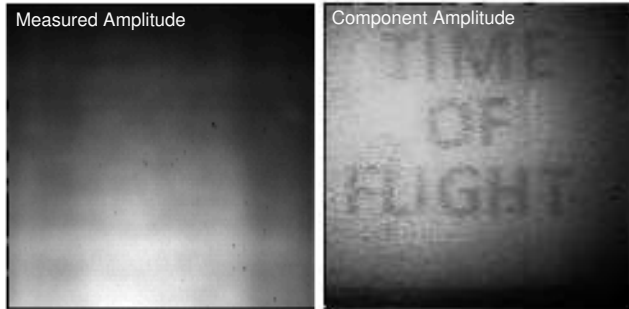

Figure 12: *We present an implementation scenario of a time of flight camera looking through a diffuser. When solving the sparse deconvolution problem and selecting the furthest return, we are able to read the hidden text.*

amplitude image that the camera observes, it is challenging to make out the text behind. However, by deconvolving and visualizing the amplitude of the Dirac from the back wall, we can read the hidden text.

4. Resolving Ranging Errors We present a simple example of resolving ranging errors. In Figure 14 we take a time of flight capture of a checkerboard grating where mixed pixels occur along edges. We show the original slice along the image as well as the corrected slice which has two discontinuous depths. While a simple TV norm would also suffice, this toy example demonstrates that deconvolution, in addition to application scenarios, can help mitigate standard multipath time of flight ranging errors.

7.2. Pitfalls

Failure cases occur when multiple light paths with a similar optical length smear into a single time slot. In Figure 15 the deconvolved slice of the corner is no better than the phase of the Fourier harmonics. Such cases remain an open problem.
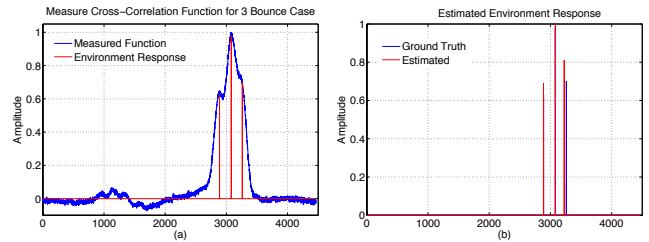

Figure 13: *Proximal matching pursuit generalizes to more bounces. 2 transparencies are placed at different depths in front of a wall, resulting in mixed pixels of three optical paths.*

Another pitfall occurs when there is an amplitude imbalance between the returns. In Figure 2, the leg on the glass unicorn in the Amplitude Image (upper-left) is specular. Very little light passes through the glass and back to the unicorn, and the measurement primarily consists of the specular return. As such, the background depth image (upper-right) still uncludes remnants of the specular leg. Similarly, in Figure 1 it is not possible to see through the specularities on the glossy vase to reveal the checkered wall behind.

8. Discussion

8.1. Comparisons

Our prototype camera combines the advantages of a sparsity based approach with custom codes. Using data acquired at a single modulation frequency we are able to obtain time profile movies of a scene at 70 picoseconds of resolution with simple hardware. Although the state-of-the art by Velten et al. obtains 2 picosecond time resolution, their approach uses laboratory grade equipment and is out of reach for most research labs. While a low-budget solution has been recently proposed by Heide et al, our method does not require special hardware for multi-frequency capture and avoids a lengthy calibration protocol. In addition, we document our achieved time resolution and observe that our approach minimizes the gap between theoretical and practical limits.
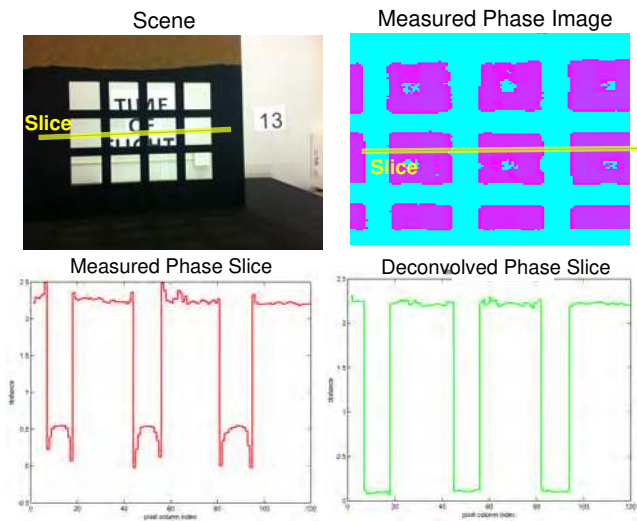
Figure 14: *A simple result, where we show that deconvolution produces a "'sharper'" depth map. The scene is a checkerboard grating in front of a wall. In the regular phase image of a checkerboard pattern, the edges have a characteristic mixed pixel effect on the edges. This is clear when plotting a horizontal slice along the checkerboard (red curve). After deconvolving, we obtain a cleaner depth discontinuity (green curve).*
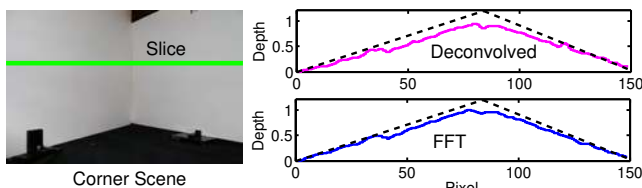

Figure 15: *Failure Case. Our results are inconclusive on a corner scene: the time profile is non-sparse and reflections smear into one time slot.*

Part of our work is focused on applications that extend time of flight technology to a broader context. Despite the increasing popularity of time of flight imaging, to our knowledge, resolving translucent objects for this modality is unexplored. In this paper, we also increase the accuracy of time of flight range measurements by removing multipath contamination. Finally, while looking through a diffuser is a well characterized problem, we have tailored it to time of flight technology.

8.2. Limitations

In our experiments, we approach the theoretical time limit of 70 picoseconds set by the FPGA clock. In the future, this limitation can be addressed by using readily available FPGA boards that have higher clock limits. The low spatial resolution of our Light Sweep Imaging technique is inherent to current time of flight cameras. State of the art time of flight cameras have four times the resolution of our prototype: we expect this to be less of a factor in coming years.

8.3. Future Improvements

Reducing the sampling period would increase the time resolution of our method. This is possible on currently available FPGA boards, including the Kintex 7 FPGA from Xilinx which supports voltage oscillations up to 2100 MHz. In addition, the direction of time of flight technology is heading toward increased modulation frequencies and higher spatial resolution, both of which can improve our time profile imaging.

8.4. Real-Time Performance

By using a single modulation frequency our technique opens up the potential for real-time performance. The method proposed by Heide et al. requires 6 hours to capture the correlation matrix: our approach–after calibration–requires only 4 seconds to capture the data required for a time profile movie of the scene. We expect that using shorter m-sequences or a compressed sensing approach to sampling time profiles would reduce the acquisition time. On the hardware side, using the on-board FPGA from Mesa and PMD cameras would cut down on our read-out times and lead to a production-quality camera.

9. Conclusion

Time of Flight cameras can be used for more than depth. The problem of multipath estimation has gained recent interest within the time of flight community, but is germane to computer vision and graphics contexts. Recent solutions to solving the multipath problem hinge upon collecting data at multiple modulation frequencies. In this paper, we show that it is possible to use only one modulation frequency when coupled with sparsity based approaches. Comparing between different time profile imaging systems is challenging; however, we offer analysis of our proposed method and find it to be on the order of 70 picoseconds in the best case. By solving the multipath problem, we have demonstrated early-stage application scenarios that may provide a foundation for future work.

References

ABRAMSON, N. 1980. Light-in-flight recording by holography. In *1980 Los Angeles Technical Symposium*, International Society for Optics and Photonics, 140–143.

BAMJI, C., ET AL., 2008. Method and system to avoid inter-system interference for phase-based time-of-flight systems, July 29. US Patent 7,405,812.

BHANDARI, A., KADAMBI, A., WHYTE, R., STREETER, L., BARSI, C., DORRINGTON, A., AND RASKAR, R. 2013. Multifrequency time of flight in the context of transient renderings. In *ACM SIGGRAPH 2013 Posters*, ACM, 46.

BRUCKSTEIN, A. M., ELAD, M., AND ZIBULEVSKY, M. 2008. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *Information Theory, IEEE Transactions on 54*, 11, 4813–4820.

BUTTGEN, B., AND SEITZ, P. 2008. Robust optical time-of-flight range imaging based on smart pixel structures. *Circuits and Systems I: Regular Papers, IEEE Transactions on 55*, 6, 1512–1525.

CARNEGIE, D. A., MCCLYMONT, J., JONGENELEN, A. P., DRAYTON, B., DORRINGTON, A. A., AND PAYNE, A. D. 2011. Design and construction of a configurable full-field range imaging system for mobile robotic applications. In *New Developments and Applications in Sensing Technology*. Springer, 133–155.

CLAERBOUT, J. F., AND F, M. 1973. Robust modelling with erratic data. *Geophysics 38*, 1, 826–844.

2008. Special issue on compressed sensing. *IEEE Signal Processing Magazine 25*, 2.

DARCHE, G. 1989. Iterative ℓ_1 deconvolution. *SEP Annual Report 61*, 281–301.

DAUBECHIES, I., DEFRISE, M., AND DE MOL, C. 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics 57*, 11, 1413–1457.

DORRINGTON, A. A., GODBAZ, J. P., CREE, M. J., PAYNE, A. D., AND STREETER, L. V. 2011. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, 786404–786404.

GODBAZ, J. P., CREE, M. J., AND DORRINGTON, A. A. 2012. Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar. In *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, 829618–829618.

GODBAZ, J. P., DORRINGTON, A. A., AND CREE, M. J. 2013. Understanding and ameliorating mixed pixels and multipath interference in amcw lidar. In *TOF Range-Imaging Cameras*. Springer, 91–116.

GUPTA, M., AGRAWAL, A., VEERARAGHAVAN, A., AND NARASIMHAN, S. G. 2013. A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision*, 1–23.

HEIDE, F., HULLIN, M. B., GREGSON, J., AND HEIDRICH, W. 2013. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics 32*, 4 (July).

HODRICK, R., AND PRESCOTT, E. 1997. Postwar u. s. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking 29*, 1–16.

JENSEN, H. W., MARSCHNER, S. R., LEVOY, M., AND HANRAHAN, P. 2001. A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM, 511–518.

KADAMBI, A., IKOMA, H., LIN, X., WETZSTEIN, G., AND RASKAR, R. 2013. Subsurface enhancement through sparse representations of multispectral direct/global decomposition. In *Computational Optical Sensing and Imaging*, Optical Society of America.

NAIK, N., ZHAO, S., VELTEN, A., RASKAR, R., AND BALA, K. 2011. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. In *ACM Transactions on Graphics (TOG)*, vol. 30, ACM, 171.

NAIK, N., BARSI, C., VELTEN, A., AND RASKAR, R. 2013. Time-resolved reconstruction of scene reflectance hidden by a diffuser. In *CLEO: Science and Innovations*, Optical Society of America.

NARASIMHAN, S. G., NAYAR, S. K., SUN, B., AND KOPPAL, S. J. 2005. Structured light in scattering media. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, IEEE, 420–427.

O'BRIEN, M. S., SINCLAIR, A. N., AND KRAMER, S. M. 1994. Recovery of a sparse spike time series by l_1 norm deconvolution. *IEEE Trans. Signal Proc. 42*, 12, 3353–3365.

RASKAR, R., AND DAVIS, J. 2008. 5d time-light transport matrix: What can we reason about scene properties. *Int. Memo07 2*.

RASKAR, R., AGRAWAL, A., AND TUMBLIN, J. 2006. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM Transactions on Graphics (TOG)*, vol. 25, ACM, 795–804.

SANTOSA, F., AND SYMES, W. W. 1986. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing 7*, 4, 1307–1330.

TAYLOR, H. L., BANKS, S. C., AND MCCOY, J. F. 1979. Deconvolution with the ell_1 norm. *Geophysics 44*, 1, 39–52.

TSIN, Y., KANG, S. B., AND SZELISKI, R. 2006. Stereo matching with linear superposition of layers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 28*, 2, 290–301.

VELTEN, A., WILLWACHER, T., GUPTA, O., VEERARAGHAVAN, A., BAWENDI, M. G., AND RASKAR, R. 2012. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications 3*, 745.

VELTEN, A., WU, D., JARABO, A., MASIA, B., BARSI, C., JOSHI, C., LAWSON, E., BAWENDI, M., GUTIERREZ, D., AND RASKAR, R. 2013. Femto-photography: Capturing and visualizing the propagation of light. *ACM Transactions on Graphics 32*, 4 (July).

WHYTE, R. Z., PAYNE, A. D., DORRINGTON, A. A., AND CREE, M. J. 2010. Multiple range imaging camera operation with minimal performance impact. In *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, 75380I–75380I.

WU, D., WETZSTEIN, G., BARSI, C., WILLWACHER, T., OTOOLE, M., NAIK, N., DAI, Q., KUTULAKOS, K., AND RASKAR, R. 2012. Frequency analysis of transient light transport with applications in bare sensor imaging. In *Computer Vision–ECCV 2012*. Springer, 542–555.

Appendix: Building a Coded ToF Camera To build your own coded ToF camera you will need 3 components:

1. A pulsed light source with 50 MHz bandwidth
2. A lock-in CMOS ToF sensor
3. A microcontroller or FPGA

The software on the microcontroller or FPGA handles the read-out from the sensor and strobes the illumination in a coded pattern. To sample the correlation waveform the FPGA software quickly shifts either the reference or illumination codes.